

INHALTSVERZEICHNIS

Automatische Erkennung von Rechnungen und Belegen	2
Das Angebot.....	2
Datenschutz	2
Probleme beim Scannen von Rechnungen	2
PDF-Rechnungen verarbeiten.....	3
Automatisierte Erfassung als Alternative	6
Preisbetrachtung	7
Hintergrund	7

AUTOMATISCHE ERKENNUNG VON RECHNUNGEN UND BELEGEN

DAS ANGEBOT

In den letzten Jahren sind eine Reihe von Anbietern am Markt erschienen, die die vollständige Automatisierung der Rechnungs- und Belegerfassung als Vorstufe für Rechnungseingangsbücher, Reisekostenerfassung und die Buchführung sowie das Banking anbieten.

Die Palette der Anbieter und Angebote ist vielfältig. Da gibt es Angebote der Banken – etwa die Deutsche Bank –, die in ihren Banking-Apps das Fotografieren und automatische Erfassen von Rechnungen anbieten, um den Rechnungsbetrag schnellstmöglich überweisen zu können. Da tritt natürlich Google auf den Plan und bietet mit Google Cloud Vision neben der Gesichtserkennung und der Bildersuche ebenfalls eine OCR (Optical Character Recognition – optische Zeichenerkennung) Funktion für Scans an. Dann gibt es Spezialanbieter wie Veryfi, Klippa oder taggun.io, die sich auf die reine Erkennung von Textinhalten in Dokumenten spezialisieren und hier vor allem Rechnungen und Belege im Blick haben – entsprechende Schnittstellen und Automatisierungen mit Buchhaltungsprogrammen werden angeboten. Und schließlich halten es einige Anbieter von Buchhaltungs-Software inzwischen für unabdingbar, den Kunden eine Vollautomatisierung der Buchhaltung anzubieten. Dank automatischer Verarbeitung von Eingangsrechnungen soll sich so die Buchhaltungsarbeit von selbst erledigen – so der Werbe-Tenor.

Warum macht INTex Publishing das nicht auch oder bietet zumindest Schnittstellen an?

DATENSCHUTZ

Zunächst stellt sich die Datenschutz-Frage – auch Rechnungen enthalten personen-bezogene Daten, zumindest dann, wenn es um Ausgangsrechnungen an private Endverbraucher geht. Aber auch bei Einkäufen Ihrer Mitarbeiter für den persönlichen Arbeitsbedarf (etwa bei Bahntickets, Hotelrechnungen etc.) sind personen-bezogene Daten enthalten. Bei einem Scan und automatisierter OCR-Bearbeitung der Rechnung geben Sie indirekt diese Daten an einen Dritten weiter, setzen also auf Auftragsdatenverarbeitung. Und hierfür müssen Sie die Vereinbarkeit des Angebots mit der DSGVO sicherstellen.

PROBLEME BEIM SCANNEN VON RECHNUNGEN

Nächster Schritt ist die Digitalisierung der Rechnungen. Bevor überhaupt an eine OCR-Verarbeitung gedacht werden kann, müssen die Belege erst mal digital vorliegen. Natürlich gibt es immer mehr Rechnungen, die man im PDF Format per Email erhält, aber weiterhin werden viele Lieferanten vom Handwerker über die Tankstelle bis hin zum Restaurant gedruckte Belege aushändigen. Diese müssen für die elektronische Weiterverarbeitung erst gescannt werden, was natürlich mit Aufwand verbunden ist. Ein Aufwand übrigens, den man sich bei der einfachen manuellen Erfassung etwa der Tankkosten ersparen kann.

Neben dem Aufwand für den Scan-Vorgang sind dann auch noch einige Hürden des Alltags zu überwinden. Die Scan-Software kann in der Regel für den OCR Vorgang nämlich nur dann ein vernünftiges Lese-Ergebnis bereitstellen, wenn die zu scannende Vorlage stimmt. Dagegen steht leider oft:

- Schief eingezogene Blätter am Scanner
- schräge Smartphone Fotos mit schlechter Belichtung und Unschärfe (bei der Arbeit mit Programmen wie OfficeLens)
- Falze, Knicke und Eselohren in der Papiervorlage
- Unsauberer Druck der Vorlage
- Kaffeeflecken
- Vergilbtes und aufgerolltes Thermopapier

Probleme der automatischen Rechnungserkennung – INTex Publishing

Auf Basis dieser Alltagsprobleme fiel ein Test, den wir mit der Deutschen Bank Smartphone Banking-App gemacht haben, sehr ernüchternd aus. Von vier getesteten Rechnungen konnte nur eine Rechnung korrekt gelesen werden, d.h. Bankverbindung, Rechnungsbetreff und Betrag waren richtig. Die korrekt erkannte Rechnung war übrigens eine Handy-Rechnung von 1&1- hier kann man vermuten, dass besondere Algorithmen für diese sehr gängige Rechnungsform zum Einsatz kamen. Bei der Rechnung eines örtlichen Handwerkers war das Testergebnis desaströs.

Hinweis: OCR Marktführer Abbyy versteckt in den Support-Dokumenten den nicht unwichtigen Hinweis, dass für gute Ergebnisse im Zweifel ein „Anlernen“ der Software an den jeweiligen Rechnungstyp notwendig sein könnte. Das lohnt sich aber natürlich wiederum nur, wenn ein Rechnungsaussteller häufiger Rechnungen verschickt.
https://help.abbyy.com/en-us/flexcapture/12/invoice_reader/ir_admintraining

PDF-RECHNUNGEN VERARBEITEN

Es ist wegen der Probleme mit den Scans vermutlich ohnehin ratsamer, die automatisierte Bearbeitung von Rechnungen auf die Rechnungen zu beschränken, die a) von vornherein digital erstellt wurden und b) als PDF-Datei vorliegen.

Wir haben hier einen Test mit der Software des Anbieters Veryfi gemacht. Zum Test dienten eine Rechnung des DSL Anbieters NetCologne, eine Rechnung des Website-Providers All-Inkl.com und eine unserer Rechnungen, die wir mit der Shop-Software Shopware erstellen.

Und das sind die Ergebnisse:

NETCOLOGNE

The screenshot shows the Veryfi website interface. At the top, there is a navigation bar with the Veryfi logo, a search icon, and links for SOLUTIONS, PRICING, BLOG, RESOURCES, SIGN IN, and a FREE TRIAL button. Below the navigation bar is a 'Try It Now' section with a dashed blue border. Inside this section, there is a cloud icon and the text 'Drag a receipt here to see a live demo'. Below this, it says 'Supported files: JPEG/JPG/PNG/GIF images and PDF documents'. The main content area shows a scanned invoice from NetCologne. The invoice header includes the NetCologne logo and contact information. The main body of the invoice is a table with columns for 'Leistungen', 'Anzahl', 'Grundpreis in EUR', 'Dauer (KMS)/ Volumen (MB)', 'Gesamtbetrag in EUR', and 'USt'. The table lists various services like 'Doppel-Flat', 'Monatliche Beträge', 'Doppel-Flat', 'Aktionsvorteil 24 Monate Rabatt auf die Grundgebühr', 'Bandbreite 100.000', and 'Mail and Cloud (ComCenter Basic)'. The total amount is 29,900 EUR. Below the table, there is a section for 'Gesamtaufstellung der Rechnungsbeträge' showing the net amount of 25,126 EUR, a 19% tax of 4,774 EUR, and a total amount of 29,900 EUR. To the right of the invoice, there is a 'Pretty | Raw JSON' view of the invoice data. This view shows a list of items with their descriptions and amounts. The items are: 'Monatliche Beträge' (EUR 0.00), 'Doppel-Flat' (EUR 29.90), 'Aktionsvorteil 24 Monate Rabatt auf die Grundgebühr' (EUR -10.00), 'Bandbreite 100.000' (EUR 10.00), 'Mail and Cloud (ComCenter Basic)' (EUR 0.00), and 'Gesamtaufstellung der Rechnungsbeträge' (EUR 0.00). The tax section shows 'Tax (19%)' (EUR 4.77), 'Subtotal' (EUR 25.13), 'Tax' (EUR 4.77), 'Tip' (EUR 0.00), and 'Total amount' (EUR 29.90). At the bottom right of the JSON view, there is a question: 'Are you satisfied with the results?' with thumbs up and down icons.

Probleme der automatischen Rechnungserkennung – INTex Publishing

Ergebnisse sind recht gut, Adresse und Rechnungsnummer sowie Rechnungsdatum und Gesamtbetrag wurden richtig erkannt.

Falsch ist die Interpretation einer Zwischenüberschrift „Monatliche Beträge“ als Positionszeile a 0.- Euro. Nicht ganz korrekt ist auch die von Netcologne eigentlich mit vier Stellen nach dem Komma ausgewiesene Netto- und Steuersumme als gerundete Beträge – dies ist etwas eigenmächtig, aber hinnehmbar.

Insgesamt ein gutes Ergebnis, das aber mit Upload und Verarbeitung ähnlich viel Zeit verschlingt, wie die manuelle Erfassung der Rechnung, sofern der Rechnungsaussteller bereits angelegt ist.

ALL-INKL.COM

The screenshot shows the ALL-INKL.COM website header with navigation links for SOLUTIONS, PRICING, BLOG, and RESOURCES, along with a search icon, SIGN IN, and a FREE TRIAL button. Below the header is a 'Try It Now' section featuring a dashed blue box with a cloud icon and the text 'Drag a receipt here to see a live demo'. Below this, it states 'Supported files: JPEG/JPG/PNG/GIF images and PDF documents'.

The screenshot displays an invoice from ALL-INKL.COM (Neue Medien Münnich) and its JSON representation. The invoice includes details such as the sender (INtex Publishing GmbH & Co. KG), the date (27.02.2020), and a table of items with columns for 'Bezeichnung', 'Menge', 'Einzelpreis', and 'Gesamtpreis'. The JSON data on the right lists fields like 'Vendor name', 'Address', 'Phone', 'Vendor type', 'Category', 'Payment', 'Invoice #', 'Date', and 'Due Date', along with a table of line items and tax information.

Bezeichnung	Menge	Einzelpreis	Gesamtpreis
Grundgebühr	1	299.40 €	299.40 €
Vorauszahlungsrabatt 10 % von 299.40 €	1	-29.94 €	-29.94 €

Line	Description	Amount
Line 1:	Grundgebühr	EUR 299.40
Line 2:	Vorauszahlungsrabatt	EUR -299.40
Tax 1:	MwSt (19%)	EUR 43.02
Subtotal:		EUR 226.44
Tax:		EUR 43.02
Tip:		EUR 0.00
Total amount:		EUR 269.46

Auch die Rechnung von All-Inkl. wird nicht wirklich schlecht erkannt.

Eigenmächtig ist die Software bei der Adresse des Rechnungsausstellers. Hier sorgt eine hinterlegte PLZ-Datenbank wohl dafür, dass aus der Rechnungsangabe 02742 Friedersdorf ein 02742 Neusalza-Spremburg wird. Das ist postalisch nicht wirklich falsch, steht aber so nicht auf der Rechnung.

Weiterhin wird die Kundennummer mit der Rechnungsnummer verwechselt.

Beim Vorauszahlungsrabatt passiert ein echter Patzer. Hier wird die Kommastelle verschoben und statt 10% Rabatt gibt es so satte 100% Rabatt.

INTEX PUBLISHING

VERYFI SOLUTIONS ▾ PRICING ▾ BLOG ▾ RESOURCES ▾ 🔍 SIGN IN FREE TRIAL

Try It Now

📄 Drag a receipt here to see a live demo
Supported files: JPEG/JPG/PNG/GIF images and PDF documents

Intex Publishing
Intex Publishing GmbH & Co. KG · Jakobstr. 9 · 53789 Esch
ULSMANN

Rechnung Nr. 20643

Siehe 1 von 1
Pos. Art.Nr. Bezeichnung Art. MwSt. Brutto Preis Brutto Gesamt

Pos.	Art.Nr.	Bezeichnung	Art.	MwSt.	Brutto Preis	Brutto Gesamt
1	HVG	Intex Hausverwaltung S	1	19 %	129,00 €	129,00 €

Gesamtbetrag Netto: 108,40 €
zuzü. 19 % MwSt: 20,60 €
Gesamtkosten: 129,00 €

Gewählte Zahlungsart: Vorkasse (Überweisung)
Gewählte Versandart: Online
Die Ware bleibt bis zur vollständigen Bezahlung unser Eigentum. Leistungsdatum entspricht Rechnungsdatum

Intex Publishing
Jahresnr. 9
53789 Esch
Tel: 02424-9172727
info@intexsoftware.de
www.intexsoftware.de
Kunden-Nr.: 011176
Rechn-Nr.: 16098
Datum: 11.03.2020

Vendor name: Intex Publishing
Address: [redacted]
Phone: [redacted]
Vendor type: [redacted]
Category: [redacted]
Payment: Other
Invoice #: 20643
Date: 2020-03-11 04:25:52
Due Date: [redacted]

Line 1: INTex Hausverwaltung S EUR 129.00
Tax 1: MwSt (19%) EUR 20.60
Subtotal: EUR 108.40
Tax: EUR 20.60
Tip: EUR 0.00
Total amount: EUR 129.00

Are you satisfied with the results? 👍 👎

Die von uns erstellte Standard-Rechnung wird – aus für uns nicht nachvollziehbaren Gründen – am schlechtesten erkannt. Obwohl unsere Adresse sogar zweifach ausgeworfen wird, ist diese überhaupt nicht erkannt worden. Automatisiert wäre also nicht klar, woher die Rechnung kommt.

Dafür stimmen diesmal die Beträge.

Fazit: Die Ergebnisse sind auf Basis digitaler PDF nicht schlecht, sogar deutlich besser als mit gescannten Druckvorlagen. Das war aber auch nicht anders zu erwarten. Fakt ist aber auch, dass keine einzige Rechnung völlig fehlerfrei erkannt worden ist. Großzügiger betrachtet war eine Rechnung korrekt, eine Rechnung fehlerhaft und eine unvollständig. Wenn das aber das Ergebnis ist, dann führt um eine manuelle Sichtung und Korrektur der automatisiert eingelesenen Rechnungen kein Weg herum. Die angedachte Effektivitätssteigerung verpufft auf diesem Wege aber schnell.

Ohnehin muss man auch klar sehen, dass die Software zwar im besten Fall eine Rechnung einlesen kann, aber eine inhaltliche Prüfung etwa auf die Frage hin, ob bestimmte Rechnungspositionen erbracht und fällig sind, kann die Software nicht leisten.

Genauso wenig kann die Automatisierung der Rechnungserfassung dafür sorgen, dass Rechnungen etwa Kostenstellen oder Projekten zugeordnet werden. Die digitale Entzifferung von Skonto-Bedingungen und Zahlungsbedingungen ist kaum möglich. Die Kennzeichnung des Zahlstatus als Offen oder Bezahlt kann ebenfalls nur durch Menschenhand erfolgen.

AUTOMATISIERTE ERFASSUNG ALS ALTERNATIVE

Und deshalb setzen wir bei INTex Publishing mehr auf die Automatisierung der manuellen Erfassung als den Versuch, den menschlichen Verstand und die Augenprüfung einer Rechnung zu ersetzen. Und es gibt eine Menge Dinge, die man zur vereinfachten Erfassung tun kann:

- Einmalige Erfassung der Rechnungsaussteller (Kreditoren). Diese können auch etwa im Excel-Format importiert werden. Somit muss bei der Erfassung von Rechnungen meist nur der Kreditor aus einer Liste ausgewählt werden – die Adresse ist dann immer korrekt und zentral korrigierbar.
- Befüllung vieler Felder mit Vorgabewerten, z.B. Datum des Rechnungseingangs
- Autofill, Auswahl-Kalender und Vorgabelisten z.B. für Zahlungskonditionen
- Automatische Wiederholung von wiederkehrenden Rechnungen etwa für Abos und Abschläge durch die Datenbank selbst
- Muster-Funktion – so kann jede Rechnung als Vorgabe für eine weitere Rechnung dienen, etwa dann, wenn ein Kreditor zwar inhaltlich unterschiedliche Rechnungen ausstellt, aber eben der Kreditor, Zahlungsziele, Skonti etc. immer gleich sind.

Nebenher kann bei der manuellen Erfassung auf die möglicherweise schwierige Erstellung eines lesbaren, sauberen Scans von Papierrechnungen verzichtet werden.

Arbeitsvorgang	Mit Automatisierung durch OCR	Ohne OCR (INTex)
Scan von Papiervorlagen	Unabdingbar in möglichst hoher Qualität	Nicht notwendig
Upload oder Weiterleitung der digitalen Rechnung	Unabdingbar für die Bearbeitung (DSGVO bedenken)	Nicht notwendig
Erfassung des Kreditors	Wird aus der Rechnung ausgelesen, u.U. immer wieder -> doppelte Datenhaltung	Einmalig zentrale Erfassung
Erfassung der Rechnungsdaten wie Kundennummer, Rechnungsnummer	Wird aus der Rechnung ausgelesen, manuelle Korrektur u.U. notwendig	Manuelle Erfassung mit Autofill Musterbuchungen
Erfassung der Skonti und Zahlungsbedingungen	Die OCR Erfassung bedeutet keine logisch-inhaltliche Erkennung	Berechnung von Skonto und Zahlungszieltag
Inhaltliche Bearbeitung und Kategorisierung	Nicht möglich -> manuelle Nacharbeit	Direkt bei der Erfassung manuell, automatisiert über Musterbuchungen
Erfassung der Rechnungssummen	Wird aus der Rechnung ausgelesen, manuelle Korrektur u.U. notwendig	Manuelle Erfassung
Erfassung der Positionszeilen	Wird aus der Rechnung ausgelesen, manuelle Korrektur u.U. notwendig	Für die Weiterverarbeitung etwa zum Bezahlen der Rechnung oder Verbuchen in der Buchführung gar nicht notwendig
Rechnungsprüfung (formal und inhaltlich)	Manuelle Nachbearbeitung	Manuell
Bearbeitungs- und Zahlstatus	Manuelle Nachbearbeitung	Manuell
Abo-Rechnungen	Genauso viel Aufwand wie normale Rechnungen	Vollautomatisierung
Ähnliche Rechnungen	Gleicher Aufwand	Muster-Rechnungen

PREISBETRACHTUNG

Nicht außer Acht lassen sollten Sie auch die Kostenfrage. Die Rechnungserfassung durch Software-Dienstleister ist nämlich ebenfalls nicht kostenlos und die Anbieter mit integrierten OCR-Features lassen sich diese gut bezahlen. Während INTex Rechnungseingang PLUS im Jahr 79.- Euro incl. MwSt. kostet, verlangt z.B. Klippa ab 40.- Euro pro Monat.

Das Einlesen einer Rechnung mit Google Vision kostet 1,50 Dollar pro Rechnung, die ersten 1000 Rechnungen sind kostenlos. Taggun dagegen verlangt 8 US-Cent pro Rechnung, kommt mit seinem Sitz in Neuseeland aber vermutlich eher nicht als DSGVO-konformer Anbieter in Betracht.

HINTERGRUND

Abschließend wollen wir Ihnen noch erläutern, warum sich Software eigentlich so schwer mit dem Lesen von Rechnungen selbst aus digital erstellten PDFs tut und das Ganze nicht einfacher geht. Bei schlechten Scans und deren digitaler Umsetzung wird man das Problem ja noch nachvollziehen können, aber bei digitalen Vorlagen?

Hintergrund ist, dass das PDF-Dateiformat in seinem Ursprung kein Datenaustauschformat im technischen Sinne ist, also nichts mit CSV-Dateien, Excel-Tabellen oder gar XML-Dateien gemein hat. Diese Dateiarten können gut importiert und exportiert werden, weil sie ein darauf optimiertes Dateischema aufweisen. PDF-Dateien aber dienen primär zur Weitergabe und Anzeige von eigentlich für den Druck vorgesehenen Daten, also von Seitenlayouts und Designs. Genaugenommen weiß das PDF-Format von den Inhalten gar nichts, es speichert primär wie und wo bestimmte Bilder, grafische Elemente und Text dargestellt werden soll. Der folgende Screenshot zeigt eine PDF-Datei in einem normalen Texteditor.



Entsprechend gibt es also keine Möglichkeit, direkt aus der PDF-Datei etwa eine Adresse oder Rechnungssumme herauszulesen, weil das PDF-Format selbst gar nicht davon weiß, dass es irgendwo derartiges gespeichert hat.

Man wird lediglich den reinen Text einer Adresse und einer Rechnungssumme in der PDF finden, kann dies aber nur mit einer Mustersuche tun. Schließlich kennt man ja vorher noch nicht die Rechnungssumme, hat keine Ahnung davon, wer die Rechnung geschrieben hat. Für die Rechnungssumme muss also nach Zahlen gesucht werden, die für die Druckausgabe vorgesehen sind. Dann muss man sich im Umfeld dieser Zahl „umsehen“, ob

irgendwo etwa „Summe“ oder „Total“ steht und kann dann die Vermutung anstellen, dass diese Zahl wohl die Gesamtsumme ist.

Üblichkeiten wie, dass die Summe wohl eher unten rechts als oben links zu finden ist, der Absender ganz oben oder ganz unten steht, werden für die Analyse ebenfalls herangezogen. All das sind aber immer nur Näherungen und diese Versuche scheitern, sobald auch nur etwas von Standards abgewichen wird. Ist der Rechnungsbogen besonders gestaltet, die „Summe“ z.B. mit „Zahlungsbetrag“ betitelt, dann wird es schnell schwierig. Die Ergebnisse der Digitalisierungs-Software spiegeln diese Probleme wider.

In den letzten Jahren sind wegen dieser Probleme Entwickler auf die Idee gekommen, dass PDF-Format in sich um einen XML Datenteil etwa für Rechnungsdaten zu ergänzen. Wenn dieser XML-Teil standardisiert ist, dann könnte man auch einfach und sicher aus diesem Bestandteil der Datei die benötigten Daten ziehen. ZugPferd und XRechnung sind Versuche, XML-Rechnungsdaten in PDFs einzubetten. INTex Rechnungseingang unterstützt das Einlesen von XRechnungen. Leider ist es nur so, dass die wenigsten digital erstellen Rechnungen über solche XML-Erweiterungen verfügen. Und ZugPferd und XRechnung sind auch nicht kompatibel.

WEITERFÜHRENDE Informationen

[PDF Format ...](#)

[XRechnung ...](#)

[ZugPferd ...](#)